Technical Disclosure Commons

Defensive Publications Series

July 03, 2017

Enabling Functionality in Content Items Based on Device Capabilities in Voice Activated Computing Systems

John D. Lanza
Foley & Lardner LLP

ENABLING FUNCTIONALITY IN CONTENT ITEMS BASED ON DEVICE CAPABILITIES IN VOICE ACTIVATED COMPUTING SYSTEMS

Voice activated computing systems provide a user with content or services in response to voice commands spoken by the user. Such systems can capture voice commands from a user, process the voice commands to determine requests and keywords in the voice commands, and provide the user with content or services related to the determined requests and keywords.

As discussed herein, a voice activated computing system processes the voice commands and selects content items to be presented to the user based, in part, on capabilities of one or more devices of the user. For example, the system can enable voice-initiated functionality for content items based on whether a user device, to which the content item is transmitted, is capable of rendering audio and accepting voice commands. The system can poll for capabilities of one or more user devices. Further, the system can determine content items that are to be presented to the user. The system, based on the polled capabilities of the user devices, can select a content item to be presented to the user device.

Figure 1 shows an example voice activated computing system. The system includes a voice assistant device, a service provider, a content provider, and a data processing system communicating over a network. The voice assistant device can be a device that accepts voice commands, and provides audio or visual output. The voice assistant can include one or more mics and cameras, such that voice commands received by the user are converted into corresponding audio signals. The voice assistant can send the audio signals to the data processing system and the service provider. The voice assistant device also can receive data such as audio signals or video signals from the data processing system or the service provider.

The voice assistant device also can include audio speakers that can convert the audio signals received from the data processing system or the service provider into sound.

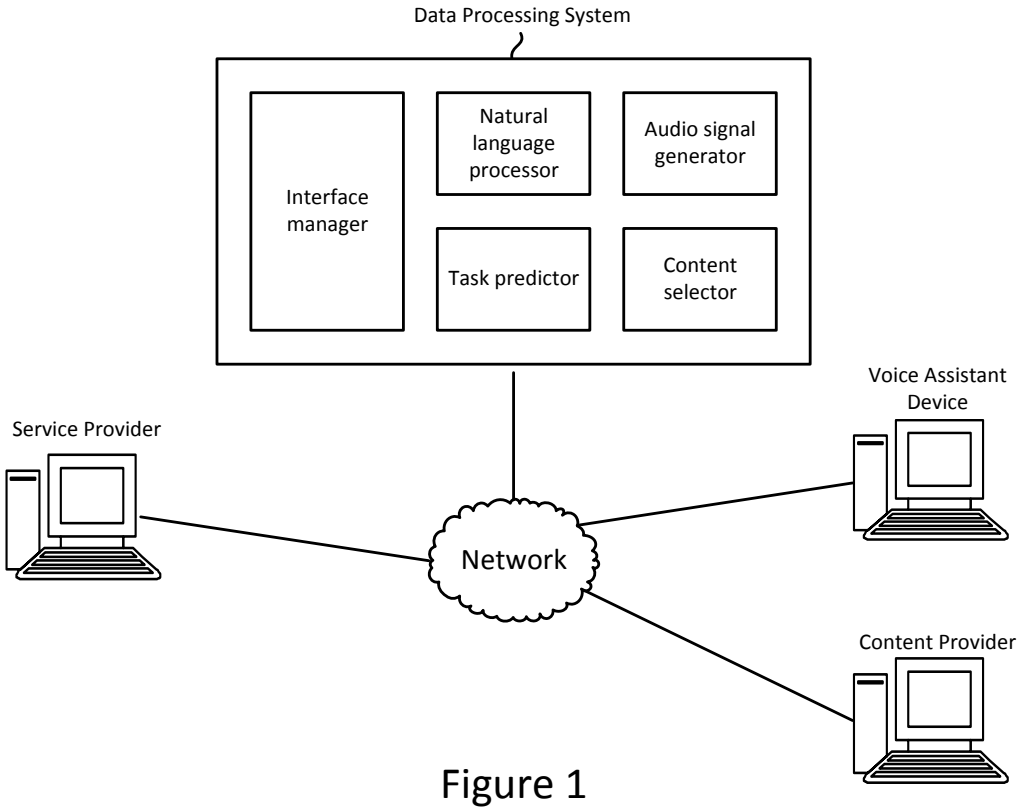

Figure 1

The data processing system can process voice commands received from the voice assistant device. The data processing system includes a natural language processor, an audio signal generator, a task predictor, an interface manager, and a content selector. The natural language processor is capable of processing voice commands included in the audio signals received from the voice assistant device. The natural language processor can convert the audio signals into recognized text by comparing the audio signals against a stored, representative set of audio waveforms, and choosing the closest matches. The representative waveforms are generated across a large set of users, and can be augmented with speech samples. After the

audio signals are converted into recognized text, the natural language processor can match the text to words that are associated, for example via training across users or through manual specification, with actions that the data processing system can serve. Basically, the natural language processor identifies requests and trigger words in the converted text, based on which the natural language processor can determine the content and actions to be carried out. The task predictor can predict tasks or actions based on the converted text, and in particular by identifying requests and trigger keywords in the converted text. The task predictor also can predict the most likely sequence in which the tasks would be executed. The content selector can select content, such as services to be offered to the user based on the actions identified by the task predictor. In addition, the content selector also can alter the sequence or the order in which the actions related to the services offered to the user are executed. The audio signal generator can generate audio signals based on the services selected by the content selector. The audio signals can be representative of voice responses or voice instructions provided to the user in response to the voice commands.

The service provider can provide one or more service to the user. For example, the service provider can be a taxi or car sharing service provider, dining or reservation service provider, and the like. The service provider can communicate with the voice assistant device independently of the data processing system and provide the user the ability to request a ride, do a dinner reservation, or avail of other services provided by the service provider. The service provider can also include a natural language processor, similar to the one discussed above in relation to the data processing system, to convert user voice commands into text, and identify requests and keywords to determine the services requested by the user.

One example voice command can include “I would like to go to a restaurant,” which the user can speak to the voice assistant device. The mics at the voice assistant device can convert the voice commands into corresponding audio signals, which are be transmitted by the voice assistant device to the data processing system over the network.

At the data processing system, the natural language processor processes the audio signal received form the voice assistant device and identify a request for a “restaurant.” In addition, the natural language processor can identify a trigger keyword “go” or “to go to,” which can indicate travel to the restaurant, such as a need for transportation to or from the restaurant. The task predictor can generate or identify actions to be performed based on the request and the trigger keyword identified in the user’s voice command. For example, the task manager can generate actions related to providing the user with a list of restaurants to choose from, making reservations at the selected restaurant, suggesting providing transportation to or from the restaurant, and the like. The task manager also can generate actions related to providing the user with sponsored content related to the identified requests or the trigger keyword. For example, the sponsored content can include deals, products, or services related to restaurants or transportation to the restaurant.

The content provider can provide both non-sponsored and sponsored content items related to the requests and trigger keywords identified in the voice command. The content items provided by the content provider can be provided to the user in response to the requests made by the user in the voice command. The content provider also can provide ads referred to by one or more service providers.

The interface manager can determine the capabilities of one or more interfaces associated with the user. For example, the interface manger can determine the capabilities of the voice

assistant device. The interface manager also can identify one or more devices associated with the user. For example, the interface manager can identify one or more devices associated with the user's account. The interface manager can poll or request the capabilities of these devices associated with the user. for example, the interface manager can determine the capabilities of these devices to receive voice commands and to render audio signals into sound. The interface manager also can determine whether the devices include display capabilities.

In providing responses to the user, the content selector can select non-sponsored and sponsored content to be delivered to the user based on the capabilities of the user device to which the content is delivered. For example, if the response requires user input, such as asking the user to select a restaurant from a list of restaurants, the content selector can select an audio signal incorporating that response and send it to the voice assistant device. However, if the user is currently in front of a device that does not have the ability to render audio, the content selector can instead select content that can be displayed to the user on a display screen of the user's device.

As another example, the content selector may include call functionality from the content delivered to the user if the user's device is capable of making phone calls. For example, in providing the user with information of a selected restaurant, the content selector can provide the user with an audio request that says: "Would you like to call the restaurant to make a reservation?" If the voice assistant device has the capability to make phone calls on behalf of the user, then the user can respond "Yes," and the voice assistant device can place the phone call to the restaurant. In another example, the content selector can provide a user's mobile device with a webpage including a button, which if pressed using an interface, calls the restaurant. If the user's mobile device allows the user to make phone calls (such as, for example, by launching a

phone application), the user can press the button to initiate the phone call. Typically, the content item can include executable code or a script that can launch a phone call application or program in response to the user responding with a voice command or by pressing a button. The phone call application is provided with the phone number of the restaurant, without the user having to enter that information.

However, if the voice assistant device or the user's mobile device do not have the capability to make a phone call, then the content selector can refrain from sending these content items to those user devices. Instead, the content selector can send the user device content that includes the information for calling the restaurant, without sending an executable code or script that can launch a phone application. The content selector can also check for other capabilities, such as the ability to launch applications such as navigation, weather, ride sharing; adequate display resolution to render the content item, the ability to provide touch interface, etc.

In another example, the user's device can include the capability to modulate the audio file received based on the type of content item being rendered. For example, the voice assistant device may use one voice tone or volume to render a list of restaurants requested by the user, and use another voice tone or volume to present the user with sponsored content, such as deals from a nearby restaurant. The change in voice tone or volume based on the content can provide aural cues to the user as to what content is being provided. In some instances, the interface manager can determine whether the user device has audio modulation capability, and select content items based on such capability.

The interface manager also can determine whether sponsored content can be sent to one or more of the user's devices other than the one via which the user sent the request. For example, the content selector may send the list of restaurants requested by the user to the voice

assistant device, and may send the sponsored content, such as deals from nearby restaurants, to the user's mobile device. In this manner, the user is not inconvenienced by having to listen to sponsored content on the voice assistant device.

Abstract

This document describes a technique for processing voice commands in a voice activated computing system. In particular, the system processes the voice commands and selects content to be sent to the user based on the capabilities of the user's device. The system can poll the user's device to determine their capabilities, such as for receiving voice commands, rendering audio responses, making phone calls, launching navigation application, accepting input via touch interface, etc. Based on the capability of a user's device, the system may select content that can be best used to provide the user with the requested information or to provide the user with sponsored content.